AN EXPERIMENT IN LITERARY AI

Can Artificial Intelligence Write a Novel Under Human Direction?

The Blue Copy Experiment

Method, literary assessment, and the question of success

Maksim Sokolov

Toronto, August 2026

MAXERGON SCHOOL

Abstract

The Blue Copy is an experimental psychological novel whose literary prose was generated entirely by artificial intelligence under close human direction. The project was conducted in Toronto from 5 to 9 August 2026. Its central constraint was simple but severe: I could identify a weakness, explain why it failed, and demand a new solution, but I could not repair the prose myself. No word, sentence, or paragraph could be manually rewritten into the novel. The finished work therefore records both the capacity of the method and the point at which further prompting ceased to improve it reliably.

This article describes the experiment, distinguishes human direction from machine execution, and offers a spoiler-free assessment of the completed English edition against practical standards of contemporary literary fiction. The verdict is qualified but affirmative. *The Blue Copy* is not evidence that an autonomous machine can produce a major novel, because autonomous authorship was never tested. It is evidence that an intensely directed human-machine process can produce, in a few days, a coherent and serious novel with sustained architecture, psychologically credible conflict, controlled motifs, and passages of genuine literary force. Its remaining weaknesses are principally editorial rather than foundational: occasional over-explanation, a tendency toward overly exact thematic correspondences, tonal uniformity, and uneven freedom in secondary characters. The experiment succeeded, but its success must be stated at the right level.

Keywords: artificial intelligence; AI-generated fiction; human-machine collaboration; experimental novel; literary criticism; authorship; prompting; contemporary fiction; Soviet historical fiction; *The Blue Copy*

1. The Question Behind the Experiment

The crude version of the question is familiar: can artificial intelligence write a novel? In that form, however, the question is too vague to be useful. A generated sequence of fifty thousand words may be called a novel because of its length while failing as fiction. Conversely, a strong work produced with machine assistance may conceal so much human rewriting that the experiment proves little about the machine's ability to sustain prose.

The Blue Copy was designed to make the question narrower and more demanding:

> *Can artificial intelligence, working under close but non-dictatorial human direction, produce all the literary prose of a serious psychological novel when the human director is forbidden to repair that prose by hand?*

This formulation matters. The experiment was not meant to discover whether a machine could produce text without human purpose, judgment, selection, or criticism. Nor was it meant to show that a prompt is equivalent to a novel. The story, characters, moral design, historical setting, philosophical questions, criteria of quality, and final editorial decisions remained human responsibilities. The machine's responsibility was equally definite: it had to turn those requirements into scenes, dialogue, narration, imagery, transitions, and finished prose, then revise its own work when criticism exposed a defect.

The result should therefore be judged as the product of a coupled creative system. Calling it wholly autonomous machine art would exaggerate the machine's independence. Calling it merely ordinary human authorship executed by a passive instrument would minimize the generative and interpretive work performed by the model. The experiment occupies the difficult space between those descriptions.

2. What Was Tested, and What Was Not

The experiment tested whether a human director and a generative model could maintain a long narrative under a strict division of labour. It tested continuity, causality, character memory, historical register, scene construction, psychological development, stylistic revision, and the ability to carry motifs across a book rather than across a single passage.

It did not test autonomous authorship. The machine did not choose the project independently, decide why the story mattered, or determine when the result was sufficient. It did not test first-draft quality, because the text passed through numerous cycles of criticism and regeneration. It did not test reproducibility in the scientific sense, because the exact number of prompts, model-computing time, and complete chat telemetry were not recorded. It did not test reader response through a blind panel, and the author of the experiment is necessarily an interested evaluator.

These limitations do not invalidate the result. They define it. The proper claim is not that an AI wrote a major novel by itself. The proper claim is that a human-directed generative system crossed a threshold that, until recently, appeared implausible: it produced an extended work that can reasonably be assessed as a novel rather than as a technological demonstration.

3. The Rules

The finished edition states two governing rules.

The first was non-intervention in the literary text. I could not insert a better word, replace an awkward phrase, or rewrite a weak paragraph. If something was false, repetitive, historically unconvincing, psychologically shallow, or stylistically mechanical, I had to describe the failure in a new prompt and require the model to solve it.

The second rule prohibited disguised dictation. Instructions had to remain at the level of story, character, causality, psychology, historical situation, rhythm, register, or general prose quality. I could explain what a passage needed to accomplish and why its present form failed, but I could not supply the desired sentence word by word.

Together, these rules created the experiment's real difficulty. They removed the easiest way to make generated prose publishable: silent human repair. They also prevented the project from becoming transcription. The text had to improve through the model's response to criticism, not through the substitution of my prose for its own.

This boundary was deliberately preserved even where I could see a local imperfection. The purpose was not to claim that every surviving sentence was ideal. It was to learn how far the chosen method could go before another kind of authorship became necessary.

4. How the Novel Was Produced

The work took place over five calendar days, from 5 to 9 August 2026, during virtually all available daytime hours. The technical record identifies the environment as ChatGPT 5.6Sol Pro and notes the use of ordinary and enhanced reasoning modes, including Pro and Ultra or Highest Intelligence, especially for literary criticism, deep revision, and final audits.

The process was divided among several sequential and parallel conversations. Manuscripts, constraints, criticisms, and editorial decisions were transferred between them. Prompts ranged from short instructions to lengthy editorial briefs. Their principal subjects included plot and causality; the architecture of the five-part novel; the internal contradictions of the characters; family relationships; moral choice and self-justification; philosophical motifs; Soviet administrative and camp environments; dialogue and historical register; letters and official documents; rhythm and syntax; the removal of repetition and artificiality; and the verification of internal consistency.

The novel was not produced by one successful command. It emerged through repeated cycles:

1. generation of a scene, chapter, or larger section;
2. close reading and identification of weaknesses;
3. analysis of causes rather than merely symptoms;
4. a new general instruction;
5. regeneration or structural revision;
6. comparison with earlier versions;
7. acceptance, rejection, or another cycle.

The model was also asked to reread its own work, detect weaknesses, compare alternative solutions, and conduct specialised audits. One documented complete editorial pass revised 593 paragraphs. Separate passes examined literary quality, historical register, philosophical references, causality, camp vocabulary, and the naturalness of the Russian and Italian prose. Russian, English, and Italian versions de-

veloped in relation to one another: a defect discovered in one language could reveal a structural problem in the others, while every version still required separate linguistic judgment.

Canonical literature served as a critical horizon, not as a template to be copied. The model was repeatedly asked to examine what the prose lacked in depth, naturalness, psychological precision, rhythm, and narrative force. The aim was not to reproduce the voice of a particular writer, but to prevent the low expectations often applied to generated text from becoming the standard of the experiment itself.

The process ended for an editorial reason. Further prompts no longer improved the manuscript reliably. Some revisions repaired one defect by damaging an earlier success; some made the prose smoother but less alive; others intensified motifs until they became too visible. The project had reached the point at which a professional prose writer's direct sentence-level intervention would probably have produced the next improvement. Crossing that line would have changed the experiment.

5. Why This Was a Difficult Literary Test

The choice of a psychological and historical novel made the task harder than the production of a conventional plot-driven narrative. *The Blue Copy* is set in the Soviet Union during the Stalinist period and its aftermath. Its central figure, Kireev, is an educated husband and father whose ideas of dignity, freedom, responsibility, and moral action have been formed by books. Life then places him in circumstances for which books provide no ready answer. The novel asks what education can give a person under pressure, what it cannot give, and how a family remembers decisions whose motives and consequences cannot be separated cleanly.

This premise requires more than factual continuity. A convincing result has to sustain several difficult things at once:

8. the protagonist must remain recognisably the same person while being changed by time and pressure;
9. moral compromise must arise from character and circumstance, not from the needs of the plot;
10. historical institutions must act through concrete procedures rather than generic menace;
11. philosophical references must influence perception and action without turning the novel into an essay;
12. family members must interpret the same person differently without any one interpretation becoming the author's final answer;
13. recurring objects and images must accumulate meaning while continuing to belong naturally to the physical world of the story.

These demands expose the typical weaknesses of generated fiction: local fluency without long memory, generic atmosphere, characters who speak with one intelligence, conflicts that resolve into an announced lesson, and symbols that appear because they are useful to the author rather than because they belong in the scene.

For that reason, the experiment cannot be evaluated merely by grammatical correctness. The relevant question is whether the completed novel achieves sustained literary organisation.

6. A Practical Standard for Comparison

There is no single official standard for contemporary fiction. For this assessment, I used a practical editorial benchmark. A serious modern novel should be evaluated on at least the following grounds:

14. narrative architecture and causal continuity;
15. depth, consistency, and development of character;
16. control of point of view and information;
17. naturalness and distinctiveness of prose;
18. dialogue that belongs to particular speakers;
19. material and historical credibility;
20. integration of image, motif, and theme;
21. emotional force that is earned by scenes rather than announced by commentary;
22. pacing, compression, and the ability to leave some meaning unstated;
23. an ending that changes the reader's understanding without merely explaining the book.

This is a higher standard than readability. It is also lower, and more useful, than comparison with a handful of canonical masterpieces. The fair question is whether *The Blue Copy* reaches the range of competent contemporary literary fiction, where its weaknesses can be discussed as literary weaknesses rather than dismissed as the inevitable output of a machine.

7. Critical Assessment of the Finished Novel

Narrative architecture and causality

The strongest achievement is structural. The English edition sustains a five-part design across thirty-six chapters and a long historical span. Earlier actions are not merely recalled; they alter the meaning of later scenes. Institutional procedures, family memories, written documents, and private acts form intersecting chains of cause and consequence. The book is constructed around moral recurrence without simply repeating the same event.

The architecture is unusually controlled for a work produced in five days. Major motifs - copies, corrections, measurements, books, tools, thresholds, official language, and the difference between a fact and the use made of it - return in altered circumstances. Their recurrence is not ornamental. Each one tests the protagonist's habit of turning experience into a system that can be understood and controlled.

The weakness is the reverse side of that control. At times the design is almost too complete. Objects return with conspicuous relevance, scenes answer earlier scenes very exactly, and a physical detail may close a moral argument with the neatness of a prepared demonstration. Contemporary literary fiction often gains power from residue, accident, and partial disorder. *The Blue Copy* sometimes allows the reader to feel the mechanism working beneath the scene.

Character and moral psychology

Kireev is the novel's most convincing achievement. He is not defined by a single virtue or failing. Discipline, intellectual seriousness, love of family, pride, physical courage, fear, vanity, and the need to make experience logically coherent operate together. The novel understands that self-justification is rarely a direct lie. More often it is an argument made from individually true statements whose arrangement conceals the conclusion toward which the speaker already wishes to move.

That psychological mechanism develops gradually and remains recognisable across changing circumstances. The protagonist's education is neither mocked nor treated as salvation. Books give him language, analogies, and standards; they also give him materials from which a persuasive defence of himself can be built. This ambiguity is sustained rather than solved.

The family is also differentiated with care. Individual characters attend to the world through different habits of mind: proof, structure, language, bodily care, memory, or practical work. These are not merely assigned traits; they become different methods of judging the same inherited past. The novel's refusal to grant one family member complete interpretive authority is a substantial strength.

Some secondary characters, however, remain more functionally designed. They are vivid in gesture, speech, or profession, but a few seem to enter principally to present a moral counterexample, embody an institutional logic, or press one aspect of the central problem. This is common in tightly constructed novels, but the best contemporary fiction gives even a minor figure the sense of a life continuing beyond the protagonist's need.

Prose, imagery, and dialogue

The prose is far above the level normally associated with an unedited machine draft. It is concrete, sensuous, and materially alert. Bureaucratic abstractions repeatedly acquire physical form through paper,

ink, stamps, worn tools, damaged clothing, food, doors, and mechanisms. The best images do two jobs at once: they establish the world and disclose a character's way of thinking.

Sentence rhythm is generally controlled. Long analytical sentences are balanced by short exchanges and physical action. Dialogue is economical, and many conversations contain pressure that is not explicitly named. The English prose is readable, serious, and largely natural. It has enough formal continuity to sustain the atmosphere of a translated or transnational historical novel without collapsing into pseudo-archaic language.

The remaining stylistic limits are visible. The narration sometimes produces one interpretive sentence too many after a scene has already made its meaning clear. Analogies are frequently exact, and the repeated use of hypothetical comparison can create a recognisable mannerism. Different characters occasionally share the same polished capacity for compressed moral observation. The prose is therefore controlled more consistently than it is singular. It has a strong house style, but not always the unpredictable pressure of a mature individual voice.

This distinction matters when comparing the novel with the best current literary fiction. The highest-level prose does not merely avoid error; it creates a voice that could not easily belong to another book. *The Blue Copy* reaches professional competence and sometimes genuine distinction, but it does not sustain that degree of linguistic inevitability on every page.

Historical and material credibility

The historical world is persuasive because power usually appears as procedure. Forms, lists, categories, permissions, reports, work norms, registration, and the vocabulary of institutions shape what characters can do. The setting is not reduced to uniforms, slogans, and fear. Labour has weight; hunger changes perception; administrative language converts uncertainty into fact; domestic life continues under pressure in small material arrangements.

This density gives the novel a lived texture. It also prevents the philosophical questions from floating free of circumstance. Moral choices occur through documents, deadlines, rooms, bodies, and consequences. That is a notable literary success.

This assessment is not a specialist archival verification. A historical novelist or historian might still identify details requiring correction. The narrower claim is that the fictional world is internally coherent, materially specific, and free of the generic historical haze common in rapid generated prose.

Emotional force

The novel earns emotion most effectively when it allows physical action to carry what characters cannot say. Family relationships are expressed through objects passed or withheld, repairs attempted, food di-

vided, papers preserved, and gestures misread. The story's emotional power does not depend solely on the extremity of its historical setting. It comes from the more ordinary difficulty of loving a person whose actions cannot be made morally simple.

Several scenes possess genuine force independent of the experiment. That is perhaps the most important evidence of success. If a passage matters only because a machine produced it, it remains a demonstration. When the reader temporarily forgets the method and attends to the human conflict, the work begins to function as fiction.

There are also moments when the novel presses too firmly for an effect. Because so many images and choices have thematic relevance, the reader receives fewer neutral surfaces than in a less controlled book. Greater restraint and compression would sometimes make the emotion stronger.

Pacing and implicitness

The overall movement is coherent, but the novel is deliberately cumulative. It returns to related moral structures from different positions, and some readers will experience that recurrence as depth while others will experience it as repetition. The middle sections occasionally pause to articulate philosophical implications that the surrounding scenes already contain.

A conventional professional edit would probably reduce some explanation, loosen a few correspondences, and allow more dialogue or action to end without a final interpretive turn. Under the rules of this experiment, however, such improvements could not be inserted manually. The surviving excess is therefore both a literary weakness and evidence about the present limit of the method.

8. Assessment at a Glance

Criterion	Assessment	Spoiler-free finding
Narrative architecture	Strong	Five-part design, durable causal links, and meaningful recurrence across a long span.
Central characterization	Strong	A complex protagonist whose virtues and weaknesses arise from the same habits of mind.
Moral psychology	Strong	Self-justification develops through credible mixtures of love, fear, reason, and pressure.
Secondary characters	Good, uneven	Often vivid and distinct, though some remain visibly functional to the design.
Prose and imagery	Good to strong	Materially precise and frequently memorable; sometimes over-explanatory or overly exact.
Dialogue	Good to strong	Economical and pressured, with occasional tonal similarity among highly articulate speakers.
Historical atmosphere	Convincing	Institutions operate through concrete language, labour, objects, and procedure.
Thematic subtlety	Mixed to strong	Themes are deeply integrated, but at times made more explicit than necessary.
Distinctive voice	Good	Coherent and controlled, though not yet singular at the level of the best literary fiction.
Publication baseline	Met, with caveats	Credible as experimental literary fiction.

9. Did the Experiment Succeed?

Yes, provided success is defined honestly.

The experiment succeeded if the question is whether artificial intelligence, under sustained human direction and a strict prohibition on manual rewriting, can generate a complete literary novel that is coherent, psychologically serious, historically textured, and worthy of criticism on ordinary literary grounds. *The Blue Copy* clears that threshold. Its central weaknesses are not the usual foundational failures of generated long fiction. The story does not dissolve into unrelated episodes; the protagonist does not reset from chapter to chapter; the themes are not pasted onto the action; and the prose is not merely grammatical filler between plot points.

The experiment did not succeed if success means autonomous machine authorship, the elimination of the human creative role, or the production of a masterpiece equal to the highest achievements of modern literature. None of those claims is supported. The human contribution was extensive and decisive: conception, direction, criticism, selection, rejection, transfer of discoveries between versions, and the decision to stop. The machine wrote all the prose, but it did not independently determine the purpose of the book.

Against competent contemporary literary fiction, the novel is respectable and in several respects strong. It meets a credible publication baseline for experimental literary fiction. Against the best contemporary novels, it remains limited by over-control: meaning is sometimes too available, motifs too obedient, and the intelligence of the prose too evenly distributed. It would be unserious to confuse rapid production with canonical achievement.

The experiment's most surprising result is therefore not perfection. It is the relocation of the problem. The question is no longer whether generated prose can remain coherent for the length of a novel. In this case, it did. The harder question is whether iterative prompting can produce the final increments of compression, idiosyncrasy, silence, disorder, and voice that distinguish competent fiction from exceptional literature.

10. What the Result Suggests About Authorship

The camera analogy used in the novel's preface remains useful, with limits. A camera performs operations that a painter's brush does not, but the artistic act can reside in framing, timing, light, selection, sequence, and the relation among images. In a directed generative project, creative agency is similarly distributed across different acts.

The analogy should not erase the model's contribution. The system did not merely record a fully formed work. It proposed scenes, language, images, transitions, and revisions that were not specified sentence by sentence. Nor should the analogy erase the human contribution. Without a sustained conception and the repeated rejection of plausible but inadequate solutions, the finished novel would not exist in its present form.

The project therefore suggests a form of authorship based less on the manual production of sentences and more on the design of constraints, the diagnosis of failure, and the selection and coordination of generated possibilities. Whether that role should be called author, director, editor, or something new is partly a legal and cultural question. Artistically, however, the division of labour can be described without mystery: the human supplied intention and judgment; the machine supplied all literary wording and participated in analysis and revision.

11. Limits of the Evidence

The experiment has several methodological limits that should remain visible.

First, it is a single project conducted by one director with one evolving system. It does not establish what another person, genre, or model would produce. Second, the process was intensive and highly iterative, so it says little about one-prompt generation. Third, the lack of exact prompt counts, machine-time telemetry, and a complete frozen log prevents strict replication. Fourth, the assessment offered here is not blind: I designed the novel and directed the experiment. Fifth, the English edition developed alongside Russian and Italian versions, so its quality cannot be attributed to a single linear generation path.

Finally, literary quality cannot be settled by technical record. A coherent structure can be demonstrated; a motif can be traced; an inconsistency can be identified. Whether a character lives in the reader's imagination, whether an image remains after the book is closed, and whether the moral uncertainty feels earned rather than manufactured depend on readers whose responses will differ.

For that reason, the strongest conclusion should remain modest. The experiment has produced evidence, not a universal law.

12. The Meaning of *The Blue Copy 1*

The finished edition calls itself *The Blue Copy 1*. The name acknowledges that the result belongs to a particular stage of generative technology and to the limits of a particular method. A later return to the same experiment might show whether a more advanced system can solve the problems that now seem resistant to prompting: excessive explanation, thematic neatness, tonal uniformity, and the difficulty of revising one local feature without flattening another.

For such a comparison to remain meaningful, the main rule should stay intact. Improvement would again have to come through machine-generated revision rather than concealed human rewriting. A future version would be most valuable not if it merely sounded smoother, but if it became less predictable in its intelligence, more selective in explanation, freer in secondary life, and more capable of leaving the reader with necessary uncertainty.

Conclusion

The Blue Copy does not answer the broad question of whether a machine can become a novelist. It answers a narrower and more practical question. Under intense human direction, with all literary prose generated by the model and with direct human rewriting forbidden, a human-machine process can now produce a sustained novel of real seriousness in a matter of days.

The result is not Dostoevsky, nor was that a reasonable threshold. It is a coherent, ambitious, and frequently accomplished experimental novel whose limitations can be named in the language of literary criticism: too much explanation in places, too visible a structure, insufficient variation of voice, and occasional pressure toward a prepared meaning. Those are significant weaknesses, but they are the weaknesses of a functioning novel.

The experiment should therefore be judged a qualified but genuine success. It surpassed the minimal test of readability, passed the harder test of sustained literary organisation, and reached a credible contemporary publication standard without proving either autonomous authorship or exceptional literary greatness. Its final value will depend on what readers discover when the novelty of the method recedes and only the characters, the language, and the unresolved moral experience remain.

Publication and Assessment Note

Primary text assessed: Maksim Sokolov, *The Blue Copy*, English edition, Maxergon School, Toronto, 2026, 199 pages. The assessment above is deliberately spoiler-free and does not describe the novel's principal decisions, outcomes, or final disclosures.

This article was prepared through human-AI collaboration under Maksim Sokolov's direction, using the completed novel, its author's note, its technical record, and the documented history of the project.